NuMI Note ANA-1093

MINOS Blind Analysis Policy

Gary Feldman
(on behalf of the Executive Committee)

1. Principles

In order to eliminate the conscious and unconscious biases that may affect physics analyses, NuMI beam oscillation analyses will be done via blind analyses.

The blinding procedure is meant only to blind the final answer while analysis procedures are being set. The minimum blinding necessary for this purpose should be used. The blinding procedure should not prevent the full exploration of the data for the purposes of calibration and verification of reconstruction code.

Before an analysis is unblinded, permission needs to be obtained, first from the cognizant working group, and then from the whole collaboration, by a procedure specified by the MAP committee.[1]

In the case of multiple analyses of the same process, the unblinding of one analysis does not bias the status of any other analyses.[2] However, for multiple analyses of the same process that are to be unblinded in the same time frame, which analysis is to be the official analysis must be determined before any are unblinded.

After unblinding, the data should be examined for errors and unforeseen backgrounds; if found these should be corrected. However, it is not permitted to otherwise modify cuts or other analysis parameters.

Each data set will be blinded only once. For further analysis, which will normally involve additional data, the old data set will return to it former blind state and the additional data will be blinded with new unknown parameters.[3]

Issues not covered in this memo will be decided by the MAP committee.

---

[1] Ideally this will also be the case for student theses. However, if a thesis advisor determines, due to time constraints on a student's thesis, that an analysis must be unblinded before the collaboration deems it ready, it will be permissible to do so, but the results will not be an official collaboration result. (See following paragraph and footnote 2.)

[2] It is anticipated that oscillation analyses will be sufficiently complex such that the casual observation of one analysis result will not bias another.

[3] This was suggested by Hugh Gallagher. It follows the principle in cryptography that a one-time pad is never reused. Footnote 2 applies here. If the analysis does not change, then the result on the old data will not change in any case. If it does change, the results of the change are not likely to be apparent.

2. Implementation

The implementation of the general blind will follow the plan devised by Nathaniel Tagg. All of the Near Detector data and out-of-spill Far Detector data will be open. The in-spill Far Detector data will be split into two streams, open and hidden, based on an algorithm with parameters chosen randomly. As specified by Nathaniel, initially the algorithm will have a constant term plus sinusoidal functions of total ADC count and event length. Random numbers determine the frequency and phase of the sinusoidal functions. Initially, each stream should contain about half of the data. Later on, we will probably want to reduce the fraction of events in the open stream. Far Detector data processed prior to the establishment of the blinding streams will remain in the open data set. The Batch Processing Group will be responsible for the verification of the blinding code.

The open stream is available for all uses. The hidden stream is usually not analyzed until it is unblinded. However, it may be analyzed for the purposes of calibration or verification of the reconstruction code when either (1) the statistics of the open sample are insufficient or (2) the blinding has biased the proposed test.[4] When testing the full data sample, individuals should avoid as much as possible viewing the relevant physics parameters of the sample, and should hide these parameters as much as possible when presenting the results of these tests.

Individuals and working groups may impose stricter blinds on the data. For example, the ν_e Analysis Working Group may wish to provide an additional blind on Far Detector ν_e candidates.

[4] An example would be an investigation of the *y* dependence of charged current events in the Far Detector. This is a possible way of verifying that the ratio between the muonic and hadronic energy scales is the same in the Near and Far Detectors. However, the full data set must be used because the blinding scheme distorts the *y* distribution.